LA-UR-21-26245

Title: Data Assimilation Methods for Dynamical Systems

Author(s): Godinez Vazquez, Humberto C.

Intended for: T-5 Summer Program

Issued: 2021-07-01

Data Assimilation Methods for Dynamical Systems

Humberto C. Godinez

T-5, Los Alamos National Laboratory

T-5 Summer Seminar
July 1st, 2021

Are you certain?

"Doubt is not a pleasant condition, but certainty is absurd."
–Voltaire
French author, humanist, rationalist, & satirist (1694 - 1778)

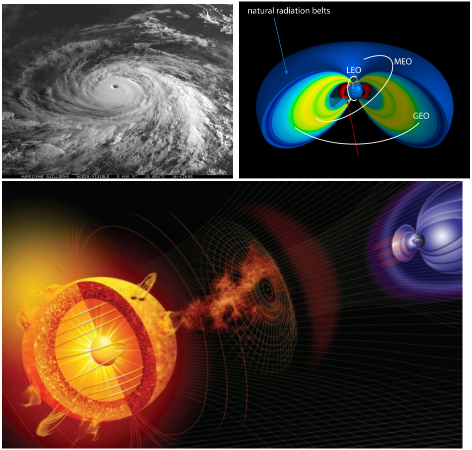
"It is in the admission of ignorance and the admission of uncertainty that there is a hope for the continuous motion of human beings in some direction that doesn't get confined, permanently blocked, as it has so many times before in various periods in the history of man."
–Richard P. Feynman

Models are not perfect
Observations are not perfect
Data assimilation combines these sources of imperfect information to provide better information.

Outline

1. Introduction to Data Assimilation
   - Types and Approaches to Data Assimilation
2. Variational Methods
3. Kalman Filter Methods
   - Kalman Filter (Linear Models)
   - Ensemble Kalman Filter
4. Lorenz 96 EnKF Example
5. Data Assimilation Examples
   - Data Assimilation for Fracture Model
   - Data Assimilation for a Global Ionosphere-Thermosphere Model
   - Ring Current Pressure Estimation with RAM-SCB using Data Assimilation

What is Data Assimilation?

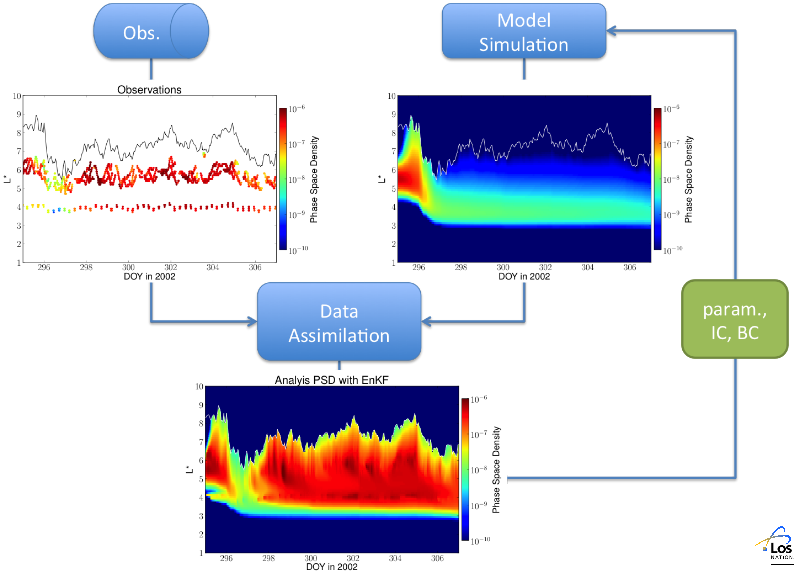


Data assimilation are methods that combine information from a model, observational data, and corresponding error statistics, to provide an estimate of the true state of a system as accurately as possible.

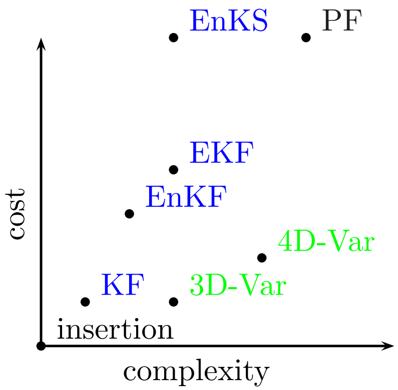
These methodologies are used in a wide range of problems, such as:

- Weather prediction
- Hurricane simulation and forecasting
- Radiation belt simulation
- Solar Physics

Data Assimilation Cycle



Type of Data Assimilation methods



Sequential Methods:

- Kalman filter (KF)
- extended Kalman filter (EKF)
- ensemble Kalman filter (EnKF)
- ensemble Kalman smoother (EnKS)
- Particle filter (PF)

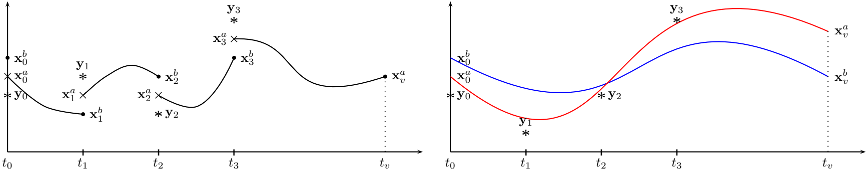
Variational Methods:

- three dimensional variational method (3D-Var)
- four dimensional variational method (4D-Var)

Hybrids:

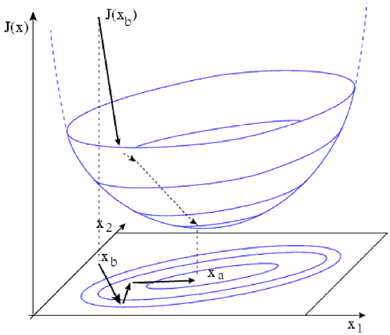
- evolving covariance matrix in 3D-Var or 4D-Var, via ensemble approximation

Approaches to Data Assimilation



Approaches:

- **Discontinuous:** assimilation individually when observations are available to compute analysis (3D-Var, Ensemble Kalman Filter).
- **Continuous:** assimilation considers all observations in a given time window to compute analysis. These methods are usually called smoothers (4D-Var, Ensemble Kalman Smoother).



The principle of variational methods is to approximate the solution (analysis) by minimizing a cost functional $\mathcal{J}$.
The solution of this minimization problem is performed iteratively with a Newton type technique (steepest decent). The analysis is the minimum of the cost function

$$\mathbf{x}^a = \underset{\mathbf{x}\in\mathbb{R}^n}{\text{argmin}}\mathcal{J}\left(\mathbf{x}\right)$$

Three- and Four-Dimensional Variational Methods (3D-Var, 4D-Var)

3D-Var

At each time-instance where observations are available, the cost function is minimized

$$\mathcal{J}(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^f\right)^T \left(\mathbf{P}^f\right)^{-1} \left(\mathbf{x} - \mathbf{x}^f\right) + \left(\mathbf{y}^o - \mathbf{H}\mathbf{x}\right)^T \mathbf{R}^{-1} \left(\mathbf{y}^o - \mathbf{H}\mathbf{x}\right)$$

where its gradient is given by

$$\nabla\mathcal{J}(\mathbf{x}) = 2\left(\mathbf{P}^f\right)^{-1} \left(\mathbf{x} - \mathbf{x}^f\right) - 2\mathbf{H}\mathbf{R}^{-1} \left(\mathbf{y}^o - \mathbf{H}\mathbf{x}\right)$$

4D-Var

4D-Var is a generalization of 3D-Var for observations that are distributed in time (smoother). The cost function then becomes

$$\mathcal{J}(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^f\right)^T \left(\mathbf{P}^f\right)^{-1} \left(\mathbf{x} - \mathbf{x}^f\right) + \sum_{k=1}^{T} \left(\mathbf{y}_k^o - \mathbf{H}_k\mathbf{x}_k\right)^T \mathbf{R}_k^{-1} \left(\mathbf{y}_k^o - \mathbf{H}_k\mathbf{x}_k\right)$$

Tangent and Adjoint are needed for both methods!!

Kalman Filter

The Kalman filter (KF) was developed by Rudolf Kalman (1960) for assimilation of linear models. The key assumptions are:

- Model is linear in time
- Both the model and the observations have a Gaussian probability distribution

Resulting analysis is estimate of true state of the system with associated uncertainty (variance).

Let $\mathbf{x}^f \in \mathbb{R}^n$, $\mathbf{M} \in \mathbb{R}^{n\times n}$

$$\mathbf{x}_i^f = \mathbf{M}\mathbf{x}_{i-1}^f + \eta_i$$

The Kalman filter analysis equations are given by

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{1}$$

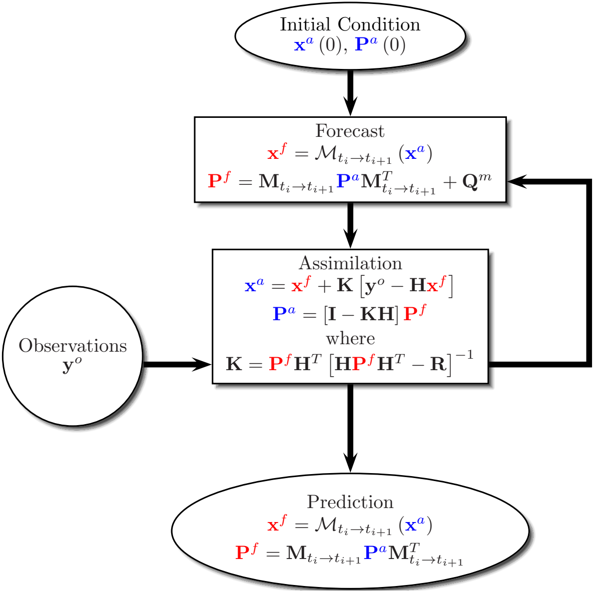
where $\mathbf{K}$ is the *Kalman Gain Matrix*, defined as

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{2}$$

The analysis error covariance matrix is given as

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f \tag{3}$$

Kalman Filter Data Assimilation Cycle



Ensemble Kalman Filter (EnKF)

The Ensemble Kalman Filter (EnKF) was first introduced by Evensen (1994) as a Monte Carlo approximation to Kalman filtering and has gained wide acceptance in data assimilation applications

- EnKF is a sequential data assimilation method that uses an ensemble of model forecast to approximate the model mean and covariance matrix
- The ensemble is updated with every analysis to reflect information provided by the observations, and is evolved using the forecast model between analysis.
- By providing a flow-dependent estimate of the model error covariance, the EnKF can optimally adjust the model forecast to newly available observations.

EnKF formulation

Let $\mathcal{M}_{t_k \to t_{k+1}}$ be the forecast model,

$$\mathbf{x}\left(t_{k+1}\right) = \mathcal{M}_{t_k \to t_{k+1}}\left(\mathbf{x}\left(t_k\right)\right) \tag{4}$$

For an vector of observations $\mathbf{y}^o \in \mathbb{R}^m$ and an ensemble of N forecast $\mathbf{x}_i^f \in \mathbb{R}^n, i = 1, \ldots, N$ the EnKF analysis equation are given by:

$$\begin{aligned} \mathbf{x}_i^a &= \mathbf{x}_i^f + \mathbf{K}\left(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f\right), \quad i = 1, \ldots, N & (5) \\ \mathbf{K} &= \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}\right)^{-1}. & (6) \end{aligned}$$

In the EnKF the forecast error covariance matrix is obtained through the ensemble of model forecast, using the relation

$$\mathbf{P}^f = \frac{1}{N-1} \sum_{i=1}^{N} \left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)^T, \tag{7}$$

where $\overline{\mathbf{x}}^f$ is the forecast ensemble average

Los Alamos NATIONAL LABORATORY EST.1943

Lorenz 96 Model Assimilation Example

The Lorenz 96 40-variable model is a toy atmospheric model which contains chaotic behavior commonly present in many atmospheric models. The Lorenz model equations are

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \left(X_{i+1} - X_{i-2}\right) X_{i-1} - X_i + F$$

where the variables $X_i, i = 1 \ldots n$ are defined in a cyclic chain such that $X_{n-k} = X_{n+k} = X_k$, and F is a specified forcing term. The system is chaotic for large values of F.

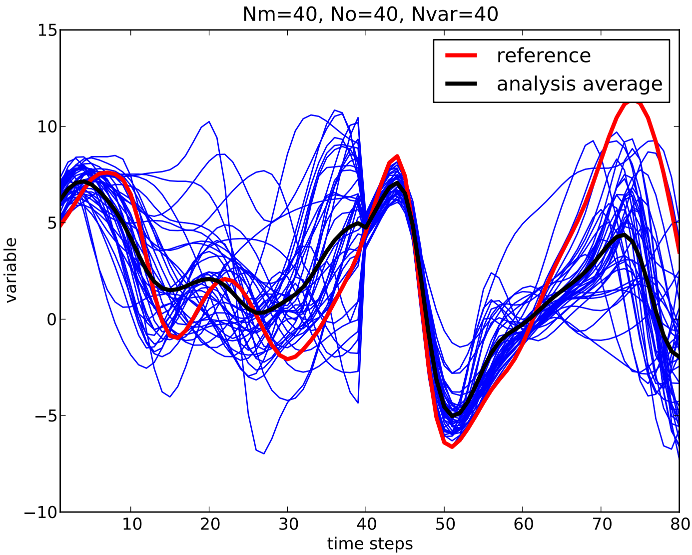
- discretized in time with a 4th order Runge-Kutta method,
- time step of $\Delta t = 0.005$ or 6 hours,
- for all experiment $F = 8.0$, and $n = 40$

Los Alamos NATIONAL LABORATORY EST.1943

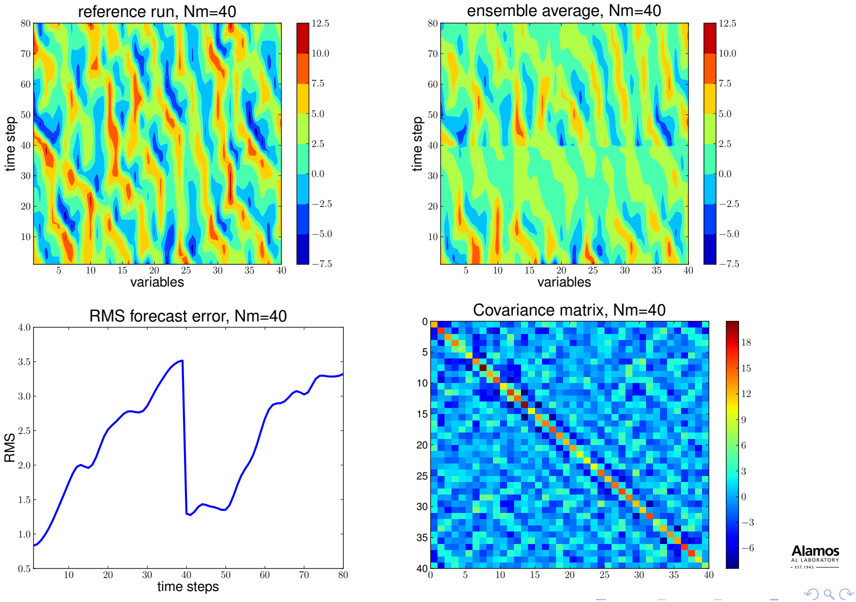
Data Assimilation Setup

- Assimilation twin-experiment, where observations are obtained from a "truth" or reference run and assimilated into a control run,
- initial condition of reference run obtained by integrating the model with 360 time-steps using $X_i = 0.0$ for $i = 2, \ldots, n$ and $X_1 = 0.001$
- initial condition for control run obtained by perturbing the reference with $N(0, 1)$,
- ensemble generated by perturbing control initial condition with $N(0, 1)$
- ensemble integrated up to 40 time-steps, with single assimilation cycle at time-step 40 and verification at time-step 80

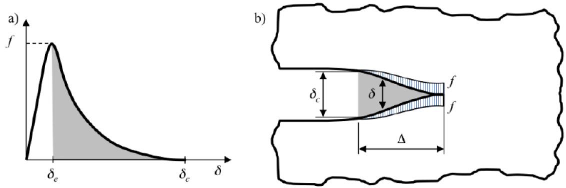
Assimilation Results with 40 members



Assimilation Results (Cont.)



Data Assimilation for Fracture Model



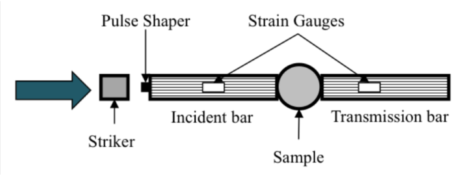
Schematics of the combined single and smeared crack model used in FDEM simulations. Figure modified from Munjiza et al. 1999

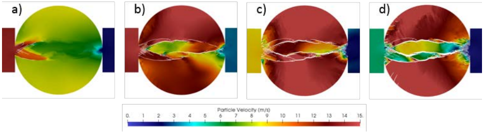
- The Hybrid Optimization Software Suite (HOSS) is a multi-physics software package, that was developed based on the combined finite-discrete element method (FDEM)
- HOSS amalgamates the finite element based analysis of continuum media with discrete element based transient dynamics, contact detection, and contact interaction solutions

Parameters of Interest:

a, b, c (softening curve), tensile strength (σ_n^{max}), shear strength (σ_t^{max}), specific energy in Mode I (normal, E_n), specific energy in Mode II (tangential, E_t)

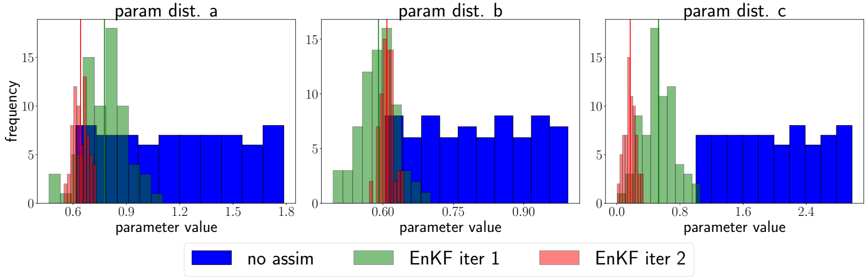
Split Hopkinson Pressure Bar (SHPB)



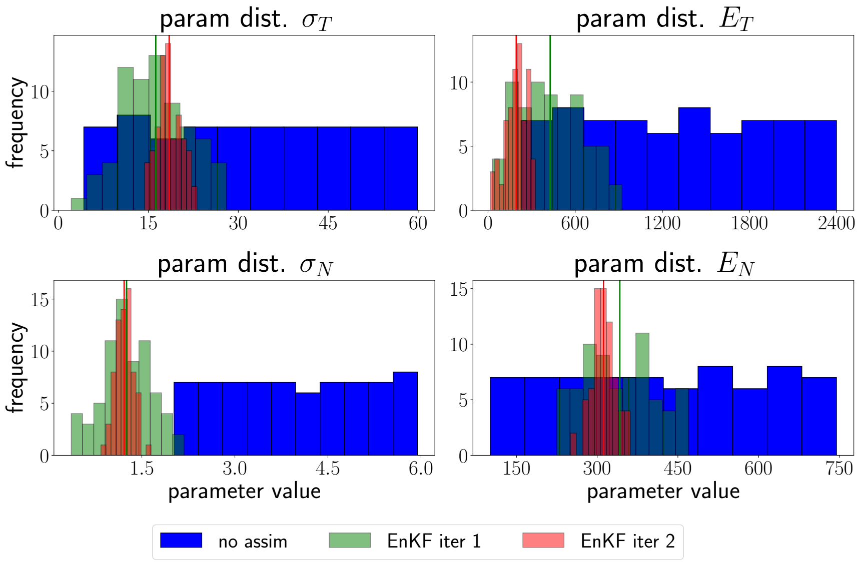


- In the SHPB experiment, a prepared test sample is placed between two bars
- Pressure pulse is generated on the left end of the incident bar by the action of a striker bar
- Pressure pulse is shaped by placing a small piece of metal on the left free end of incidence bar
- Striker travels at certain velocity and hits the incident bar

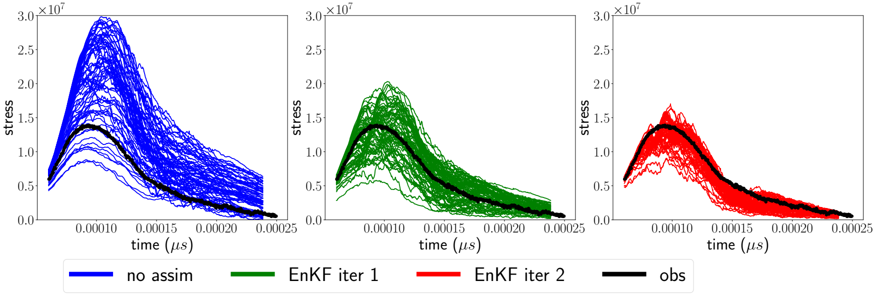
Assimilation Results



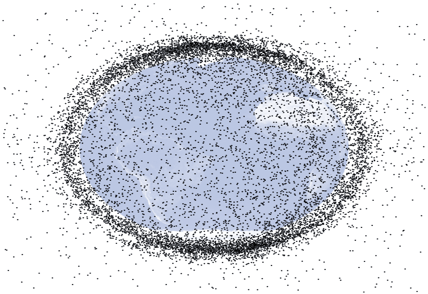
Assimilation Results (Cont.)

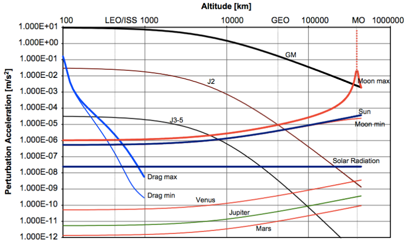


Assimilation Results (Cont.)



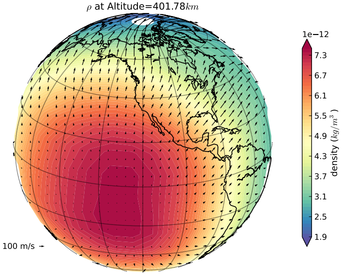
Data Assimilation for a Global Ionosphere-Thermosphere Model

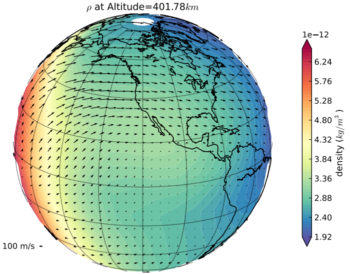




- Space debris poses a danger to satellite infrastructure, need accurate orbit prediction
- **Key element is specification of ionosphere-thermosphere environment**
- Physics-based models posses the potential to provide an accurate prediction
  - Global Ionosphere-Thermosphere Model (GITM)
  - TIEGCM
- **Data assimilation is a tool to calibrate and correct the density model output by using observational data**

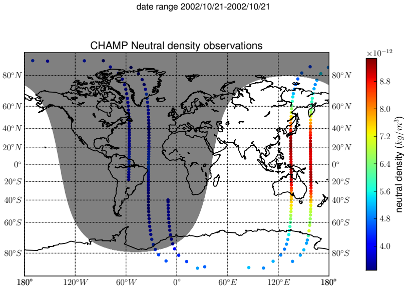
The Global Inosphere-Thermosphere Model (GITM)

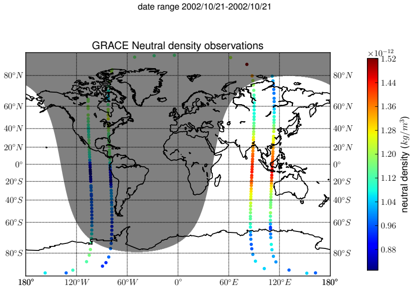




- GITM: physics based model that solves the full Navier-Stokes equations for density, velocity, and temperature for a number of neutral and charged components
- explicit solves for the neutral densities of O, O_2, $N\left({}^2D\right)$, $N\left({}^2T\right)$, $N\left({}^4S\right)$, N_2, NO, H, and He; and the ion species $O^+\left({}^4S\right)$, $O^+\left({}^2D\right)$, $O^+\left({}^2P\right)$, O_2^+, N^+, N_2^+, NO^+, H^+, and He^+,
- Solves non-hydrostatic thermosphere
- main drives is solar UV radiation (measured by $F_{10.7}$

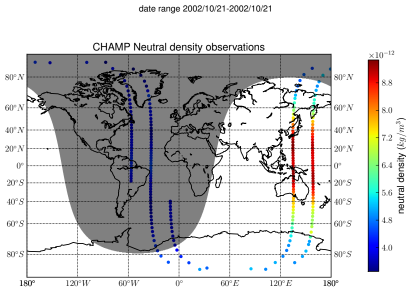
CHAMP and GRACE





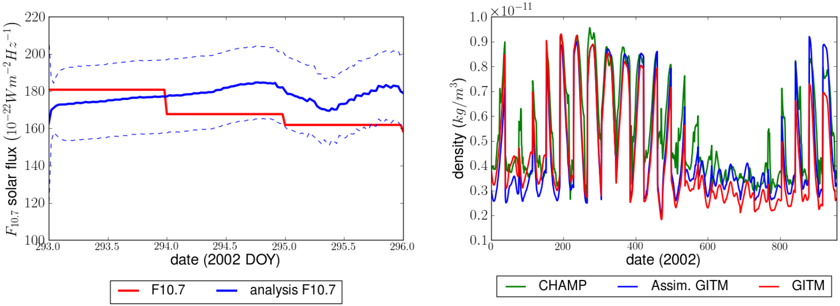
- CHAMP and GRACE provide neutral density derived observations
- use CHAMP observations for assimilation with LETKF

Data Assimilation CHAMP



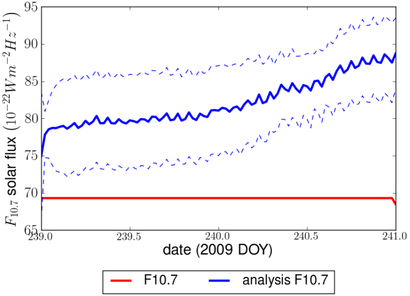
- observation taken from CHAMP for Oct 21 2002 to Oct 31 2002
- assimilation done with 10 ensemble member, using LETKF
- observations assimilated every 30 minutes
- local region set to 200 km
- “calibrated” $F_{10.7}$ parameter
- ensemble generated with mean $F_{10.7} = 165.89$ and standard deviation of 10.0

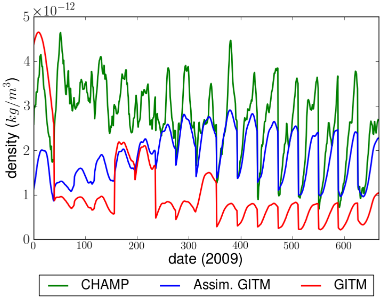
Assimilation during Solar Maximum



- Assimilation performed for October 21–24, 2002, during solar max
- GITM provides an accurate estimate for the ionosphere-thermosphere
- assimilated $F_{10.7}$ oscillating closely to measured $F_{10.7}$, not much correction is needed for GITM match observed density from CHAMP

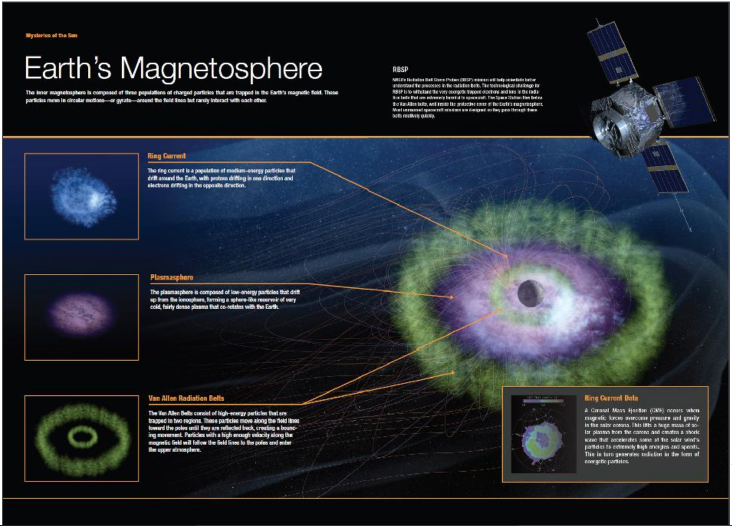
Assimilation during Solar Minimum





- Assimilation performed for August 28–31, 2009, during solar minimum.
- During solar minimum, more complex internal processes dominate ionosphere-thermosphere, GITM unable to provide accurate representation
- Assimilation provides significant changes to $F_{10.7}$, GITM seems to get closer to observed density from CHAMP

Print Close


Mysteries of the Sun
Earth's Magnetosphere
The inner magnetosphere is composed of three populations of charged particles that are trapped in the Earth's magnetic field. These particles move in circular motions—or gyrate—around the field lines but rarely interact with each other.
RBSP
NASA's Radiation Belt Storm Probes (RBSP) mission will help scientists better understand the processes in the radiation belts. The technological challenge for RBSP is to withstand the very energetic trapped electrons and ions in the radiation belts that are extremely harmful to spacecraft. The Space Station flies below the Van Allen belts, well inside the protective cover of the Earth's magnetosphere. Most unmanned spacecraft missions are designed so they pass through these belts relatively quickly.
Ring Current
The ring current is a population of medium-energy particles that drift around the Earth, with protons drifting in one direction and electrons drifting in the opposite direction.
Plasmasphere
The plasmasphere is composed of low-energy particles that drift up from the ionosphere, forming a sphere-like reservoir of very cold, fairly dense plasma that co-rotates with the Earth.
Van Allen Radiation Belts
The Van Allen Belts consist of high-energy particles that are trapped in two regions. These particles move along the field lines toward the poles until they are reflected back, creating a bouncing movement. Particles with a high enough velocity along the magnetic field will follow the field lines to the poles and enter the upper atmosphere.
Ring Current Data
A Coronal Mass Ejection (CME) occurs when magnetic forces overcome pressure and gravity in the solar corona. This lifts a huge mass of solar plasma from the corona and creates a shock wave that accelerates some of the solar wind's particles to extremely high energies and speeds. This in turn generates radiation in the form of energetic particles.

Ring Current Pressure Estimation with RAM-SCB using Data Assimilation

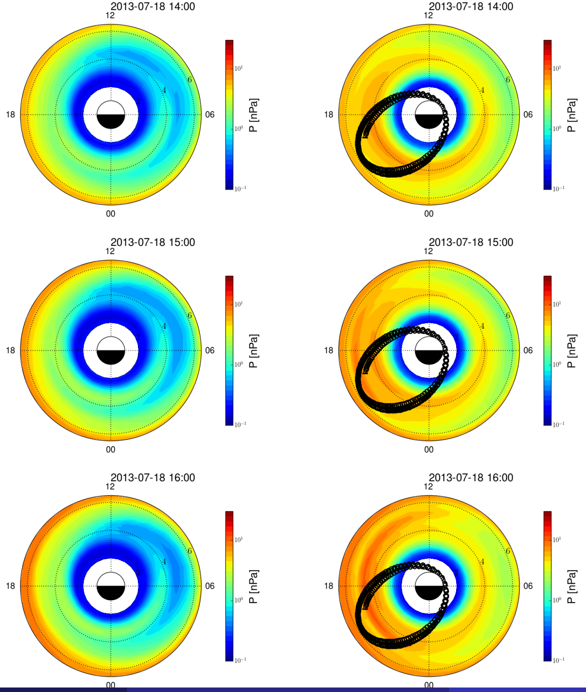
- Capturing plasma injections on the inner magnetosphere is important for understanding formation/evolution of the ring current.
- Use Data Assimilation to estimate the ring current using the Ring Current-Atmosphere Interactions Model with Self-Consistent Magnetic field (RAM-SCB)
- Use flux data from the Van Allen Probes.

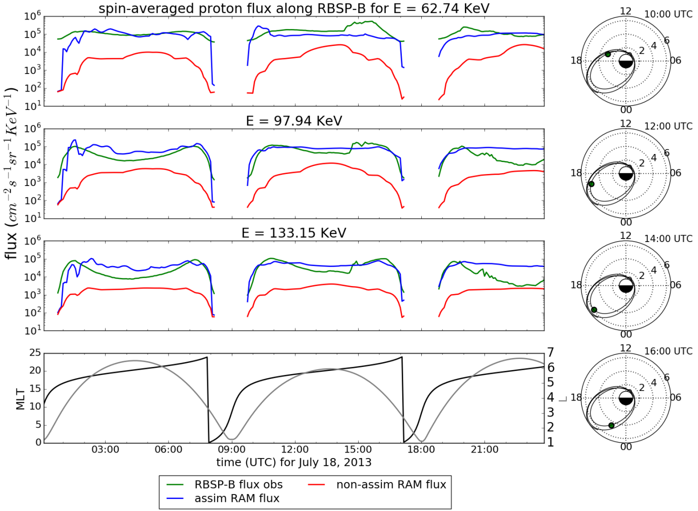
RAM-SCB Model

- physics-based model used to simulate ring current dynamics
- The RAM-SCB computes particle phase space distributions for ions and electrons on the equator inside the geosynchronous orbit for different pitch angles and energies in prescribed electric and magnetic fields

Assimilation Setup

- 20 ensemble simulations are used, generated by perturbing initial flux in RAM-SCB and running the model forward
- Assimilate proton flux data from the Van Allen B probe, validate with Van Allen A probe


2013-07-18 14:00
2013-07-18 14:00
2013-07-18 15:00
2013-07-18 15:00
2013-07-18 16:00
2013-07-18 16:00
12
18
06
00
P [nPa]


spin-averaged proton flux along RBSP-B for E = 62.74 KeV
E = 97.94 KeV
E = 133.15 KeV
flux $(cm^{-2}s^{-1}sr^{-1}KeV^{-1})$
MLT
L
time (UTC) for July 18, 2013
03:00
06:00
09:00
12:00
15:00
18:00
21:00
RBSP-B flux obs
assim RAM flux
non-assim RAM flux
10:00 UTC
12:00 UTC
14:00 UTC
16:00 UTC

Summary

- Wide range of assimilation methodologies, most are easy to implement and effective.
- Data assimilation is an active area of research (both in theory and applications).
- Assimilation in space physics used for gap filling, forecast, nowcast and/or reanalysis, reduction and specification of uncertainties
- Data assimilation offers paradigm shift from previous working relation between modelers and experimentalist, work more closely together.

Challenges ahead

- Incomplete physics $\rightarrow$ models error
- Uncertainties in observation operator
- Sparse observational field
- Non-Gaussian distributions for models and observations

References

Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53:343–367.

Godinez, H. and Koller, J. (2012). Localized Adaptive Inflation in Ensemble Data Assimilation for a Radiation Belt Model. *Space Weather*, 10:S08001.

Godinez, H., Lawrence, E., Higdon, D., Ridley, A., Koller, J., and Klimenko, A. (2015). Specification of the Ionosphere-Thermosphere Using the Ensemble Kalman Filter. *Lecture Notes in Computer Science*, 8464:274–283.

Godinez, H., Yu, Y., Lawrence, E., Henderson, M., Larsen, B., and Jordanova, V. (2016). Ring current pressure estimation with ram-scb using data assimilation and van allen probe flux data. *Geophys. Res. Lett.*, 43.

Hickmann, K. S., Godinez, H. C., Henney, C. J., and Arge, C. N. (2015). Data assimilation in the ADAPT photospheric flux transport model. *Solar Physics*, 290(4):1105–1118.

Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press.